Semiparametric Tail Index Regression

RUI LI
School of Statistics and Information, Shanghai University of International Business and Economics

CHENLEI LENG*
Department of Statistics, University of Warwick

JINHONG YOU*
School of Statistics and Mathematics, Shanghai Lixin University of Accounting and Finance

Abstract. Understanding why extreme events occur is often of major scientific interest in many fields. The occurrence of these events naturally depends on explanatory variables, but there is a severe lack of flexible models with asymptotic theory for understanding this dependence, especially when variables can affect the outcome nonlinearly. This paper proposes a novel semiparametric tail index regression model (STIR) to fill the gap for this purpose. We construct consistent estimators for both parametric and nonparametric components of the model, establish the corresponding asymptotic normality properties for these components that can be applied for further inference, and illustrate the usefulness of the model via extensive Monte Carlo simulation and the analysis of return on equity data and Alps meteorology data.

Key words. Asymptotic normality; Extremes; Pareto-type distribution; Semiparametric regression; Tail index.

Running Head. Semiparametric Tail Index Regression

1 Introduction

The study of the tail behavior of random events is of major scientific interest in many applied fields, ranging from structural engineering to finance, meteorology, earth sciences, traffic prediction, geological engineering, and so on. To this end, extreme value theory has been developed and widely applied to model extreme events; See, for example, Beirlant et al. (2004) for a comprehensive review.

* Corresponding authors: Leng C., Email: C.Leng@warwick.ac.uk; You J., Email: johnyou07@gmail.com.

A central topic to analyze extremes statistically is how to efficiently estimate the tail index that is directly related to the tail behavior of random events in which a low tail index corresponds to a high probability of extreme events. Since the pioneer work of Hill (1975) on tail index estimation with heavy-tailed distribution, many methods have been developed. Davison and Smith (1990) considered the estimation of the tail index under high threshold cases. Beirlant et al. (1996) employed the right tail of a quantile map under Pareto distribution to estimate the tail index. Davison and Ramesh (2000) proposed a semiparametric smoothing estimating approach by fitting a generalized extremum distribution with local polynomials. Danielssona et al. (2001) proposed a two-step Bootstrap sampling procedure to identify the number of extreme order statistics and estimate the tail index. Beirlant et al. (2006) established the local asymptotic normal property for the estimated tail index in Pareto and Weibull distributions in a semiparametric model. Gomes et al. (2008) proposed a weighted estimator for the positive tail index in a strict Pareto model and studied the deviation of its asymptotic distribution. Müller and Rufibach (2009) suggested that the smoothing empirical distribution of a log-concave density and conditional density can improve the efficiency for estimating the tail index. Gabaix and Ibragimov (2011) used Rank-1/2 least squares method to alleviate the bias of traditional tail index estimation for small sample cases. Beran and Schell (2012) established consistent estimation with bounded bias and robustness in small sample via optimizing Pareto maximum likelihood function. Zhang et al. (2013) proposed tail dependent regression using Logistic transformation on linear combination of covariates and studied asymptotic properties by maximizing approximated likelihood. Boucheron and Thomas (2015) proposed an adaptive Hill estimator using Lespki model selection procedure and derived the oracle inequality and lower bound accordingly. McElroy and Nagaraja (2016) studied the estimation of the tail index with fixed tuning parameters. Nemeth and Zempleni (2017) developed an approach for estimating the high tail index that also improves Hill estimator in the small sample scenario. Jia et al. (2018) proposed a semiparametric tail index estimation in which the nonparametric component controls the bias.

Despite the progress made in the aforementioned works, they often assume that the tail index is a constant independent of explanatory variables. In practice however, it is very natural to expect that covariates may play an essential role in leading to extreme events.

Assuming a linear dependence, Wang and Tsai (2009) developed a parametric tail index model, while Nicolau and Rodrigues (2019) proposed a regression-based estimator that does not involve order statistics. For inference, Ma et al. (2018) applied empirical likelihood to construct confidence regions for the regression coefficients in their tail index model. In many applications, the relationship between the tail index and covariates may be very complicated and the simple linear structure, though still useful as a crude approximation, is incapable of modeling the possible fluctuation of the tail index particular due to, for example, critical social events and macro/micro-policies. As a concrete example, foreign exchange markets in particular have been characterized by turbulence and volatility with most exchange rates featured by extreme variations and strong instability which carry significant repercussion effects for international trade, asset prices and other variables (Ibragimov et al., 2013; Straetmans and Candelon, 2013). In addition, many empirical works have shown that the tail index is often affected by variables nonlinearly. Towards this, Chavez and Davison (2005) employed the additive model to analyze sample extremes using spline smoothing and penalized likelihood, but did not study the asymptotic theory of their estimation systematically. Motivated by the discussion above, we aim to develop a more adaptive and flexible semiparametric model in this article to characterize such nonlinear dependence and establish the large sample theory for the resultant estimators.

Semiparametric models have been well established for modelling independent data (Chen, 1988; Speckman, 1988; Härdle et al., 2012), longitudinal data (Lin and Carroll, 2001; He et al., 2002), and missing data including (Liang et al., 2004), but not for extreme events. The main contribution of this paper is to apply these models to study the tail index of extreme events and rigorously establish the corresponding asymptotic properties, thus substantially generalizing the parametric approach proposed in Wang and Tsai (2009). We emphasize that establishing the theory for our model is much more challenging than the model in Wang and Tsai (2009) as the estimation of parametric and nonparametric components must be carefully analyzed to incorporate the generality of a more flexible tail index model, and more challenging than the aforementioned semiparametric models since in studying extreme events, a sample selection with an unknown sample size must be carried out to deal with the extreme events but not the whole sample. Operationally, the nonlinearity of the tail index function necessitates an iterative algorithm that involves the choice of tuning parameters based on subsets of data with different sample sizes, making the theoretical analysis very difficult. In addition, the interaction of parameters and complex technical details lead to heavy computation burden and a dilemma in studying asymptotic theory.

To compute the estimator, we propose an easy-to-implement algorithm for the parametric and nonparametric parts. We then establish their asymptotic normality accordingly in which the estimated variance can be applied for statistical inference. This variance reflects the possible volatility of the tail index in some sense based on the partially linear structure of the tail index function. More specifically, as an example, in the return on equity (ROE) data analysis in Section 5.2 , the oscillation of covariates including leverage ratio, the uncertainty of financial assets and its turnover ratio often lead to return-risks for many enterprises, and such risk is just attributed to the volatility of the tail index intuitively. Further, we use the proposed method to analyze Alps meteorology and find significant features causing extreme climate. It is worth mentioning that in both applications, some covariates indeed exert significant nonlinear effects on their response variables that can be observed from the fitted curves and confidence bands in Figure 4.

The remainder of this article is organized as follows. In Section 2, we introduce the semiparametric tail index model and propose an iteratively fitting approach to estimate the parametric and nonparametric components via local linear smoothing with the quasi-likelihood approach. In Section 3, we establish the asymptotic normality for both the estimated nonparametric and parametric components under some regular conditions. In Section 4, we discuss the choice of tuning parameters and summarize the iterative algorithm in estimation. Numerical simulations and two real data sets are presented in Section 5 to illustrate the performance of our proposed method. We conclude the paper in Section 6. All the technical details are provided in the Appendix.

2 Model and method

Let $(Y_i, \mathbf{X}_i, Z_i)_{i=1}^n$ be independent observations of random variables $(Y, \mathbf{X}, Z)$, where $Y \in \mathbb{R}$ is the response variable of interest and $(\mathbf{X}, Z)$ are covariates with $\mathbf{X} \in \mathcal{X} \subset \mathbb{R}^p$ and, without loss of generality, $Z \in \mathcal{Z} \subset \mathbb{R}$. In addition, let $F(y; \mathbf{x}, z) = pr(Y \leq y \mid \mathbf{X} = \mathbf{x}, Z = z)$ be the cumulative distribution function of Y conditional on $(\mathbf{X}, Z)$. In this paper, we use

$S(y; \mathbf{x}, z) = 1 - F(y; \mathbf{x}, z)$ to denote the conditional survival function and study the following Pareto-type model

$$S(y; \mathbf{x}, z) = y^{-\alpha(\mathbf{x},z)} L(y; \mathbf{x}, z), \tag{1}$$

where $\alpha(\mathbf{x}, z)$ is an unknown tail index function that characterizes the dependence of the tail behavior of Y on $\mathbf{X}$ and Z. In this model, $L(y; \mathbf{x}, z)$ is assumed to be a slowly varying function in the sense that $L(yt; \mathbf{x}, z)/L(y; \mathbf{x}, z) \to 1$ when $y \to \infty$ for some constant $t > 0$, and relies on $(\mathbf{x}, z)$ in the form of

$$L(y; \mathbf{x}, z) = c_0(\mathbf{x}, z) + c_1(\mathbf{x}, z) y^{-\beta(\mathbf{x},z)} + o(y^{-\beta(\mathbf{x},z)}),$$

where $c_0(\mathbf{x}, z), c_1(\mathbf{x}, z)$, and $\beta(\mathbf{x}, z)$ are unknown functions of $\mathbf{x}$ and z. Moreover, $c_0(\mathbf{x}, z)$ and $\beta(\mathbf{x}, z)$ are assumed to be uniformly bounded below from zero, and the term $o(y^{-\beta(\mathbf{x},z)})$ is a remainder dominated by $y^{-\beta(\mathbf{x},z)}$ for each pair $(\mathbf{x}, z)$.

For $\alpha(\mathbf{x}, z)$, we propose the following semiparametric tail index regression (STIR) model by denoting $\alpha(\mathbf{x}, z)$ on the logarithmic scale as

$$\log(\alpha(\mathbf{x}, z)) = \mathbf{x}^\top \boldsymbol{\theta} + \eta(z)$$

with an unknown parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$ and an unknown univariate smooth function $\eta(\cdot)$. The proposed tail index function $\alpha(\mathbf{x}, z)$ is evidently more flexible than the parametric structure of $\log(\alpha(\mathbf{x})) = \mathbf{x}^\top \boldsymbol{\theta}$ in Wang and Tsai (2009). In this paper, we focus on modelling the function $\eta(z)$ on a univariate variable z. Extension to model multiple variables in the nonparametric component is straightforward, as we will elaborate later, although care must be taken to deal with the issue of curse of dimensionality. Model (1) indicates that when y approaches infinity, a smaller tail index $\alpha(\mathbf{x}, z)$ corresponds to a slower rate at which $S(y; \mathbf{x}, z)$ decays to zero, giving rise to larger likelihood of having extreme values in Y.

Model (1) is very general in the sense that it removes the boundedness assumption on the response variable (Davison and Smith, 1990) and includes many useful heavy-tailed distributions such as Pareto, generalized Pareto, t and F distributions, as special cases without specifying the exact form of $L(y; \mathbf{x}, z)$. Particularly, when the tail index is free of covariates, model (1) has appeared in Hall (1982) and Beirlant et al. (2004) among others. For the Pareto-type model, we can write the survival function for a single observation $(Y_i, \mathbf{X}_i, Z_i)$ as

$$S(Y_i; \mathbf{X}_i, Z_i) = Y_i^{-\alpha(\mathbf{X}_i, Z_i)} \left\{ c_0(\mathbf{X}_i, Z_i) + c_1(\mathbf{X}_i, Z_i) Y_i^{-\beta(\mathbf{X}_i, Z_i)} + o(Y_i^{-\beta(\mathbf{X}_i, Z_i)}) \right\},$$

in which the approximated probability density function of Y_i conditional on $(\mathbf{X}_i, Z_i)$ and $Y_i > w_n$ can be seen as

$$f(y; \mathbf{x}, z) = \alpha(\mathbf{x}, z)(y/w_n)^{-\alpha(\mathbf{x}, z)} y^{-1}. \tag{2}$$

In this expression, we have used a tuning parameter $w_n > 0$ to control the sample fraction. Here and after, we denote $n_0 = \sharp\{i : Y_i > w_n\}$ as the effective sample size used in estimation. This is reminiscent of the left truncation technique employed in survival analysis that can circumvent the random censoring issues commonly encountered (Tsai, 1988; Hudgens, 2005; Kalbfleisch and Prentice, 2011; Chen, 2019). More specifically, left truncation has been applied to failure time modeling in lifetime data analysis, see, for example, Cai et al. (2007, 2008) that considered a semiparametric hazard model without distributional assumptions. When the survival distribution is assumed, Li and Lee (2011) studied the Cox-proportional hazards model based on threshold regression and inverse Gaussian distribution. Shao et al. (2014) applied a varying coefficient model to analyze interval censored data with a cured proportion in medical studies under a location-scale family of distributions. Xia et al. (2016) considered covariate-adjusted screening and variable selection in an accelerated failure time model, among others. It's worth mentioning that the left truncation procedure may often induce selection bias and adversely affect estimating precision and inference, especially when the effective sample size is small (Howards et al., 2006; Cain et al., 2011; Schisterman et al., 2013).

For estimation, we apply the results in Severini and Staniswalis (1994) to the probability density function in (2) to obtain the following quasi-likelihood function

$$\mathcal{L}_n(\boldsymbol{\theta}, \eta) = \sum_{i=1}^{n} \left\{ \log(Y_i/w_n) \exp(\mathbf{X}_i^\top \boldsymbol{\theta} + \eta(Z_i)) - \mathbf{X}_i^\top \boldsymbol{\theta} - \eta(Z_i) \right\} I(Y_i > w_n). \tag{3}$$

To estimate the function $\eta(z)$ in (3), we adopt local linear smoothing in Fan and Gijbels

(1996) by approximating $\eta(Z_i)$ via the first-order Taylor expansion at some z satisfying

$$\eta(Z_i) = \eta(z) + \eta^{(1)}(z)(Z_i - z) + O_p(h_n^2) \tag{4}$$

for $\mid Z_i - z \mid \leq h_n$, where $h_n > 0$ is a bandwidth approaching zero as the sample size $n \to \infty$. The choice of bandwidth h_n is important in local smoothing, particularly for the profile estimation of the parametric component in semiparametric regression since undersmoothing is often employed to guarantee the $\sqrt{n}$-consistency of the parametric estimator. In this article, we apply undersmoothing in our first-stage estimation in equation (6) below to derive the asymptotic normality of estimated parameters. More specific discussion on the bandwidth h_n is presented in Section 4.1.

Substituting (4) into (3) and removing the redundant components of order $O(n_0 h_n^2)$ based on assumptions (C4) and (C5) in Section 3, we derive the estimators of $\boldsymbol{\theta}$ and $\eta(\cdot)$ by minimizing

$$\sum_{i=1}^{n} K_{h_n}(Z_i - z) \left\{ \log(Y_i/w_n) \exp(\mathbf{X}_i^\top \boldsymbol{\theta} + \bar{\mathbf{Z}}_i^\top \boldsymbol{\delta}) - \mathbf{X}_i^\top \boldsymbol{\theta} - \bar{\mathbf{Z}}_i^\top \boldsymbol{\delta} \right\} I(Y_i > w_n), \tag{5}$$

where $K_{h_n}(\cdot) = K(\cdot/h_n)/h_n$ is a nonnegative and symmetric kernel function, $\bar{\mathbf{Z}}_i = (1, (Z_i - z)/h_n)^\top$ and $\boldsymbol{\delta} = (\delta_0, \delta_1)^\top = (\eta(z), \eta^{(1)}(z))^\top$. For any given $\boldsymbol{\theta}$, we take the first partial derivative of (5) with respect to $\boldsymbol{\delta}$ and solve the following estimating equation

$$\frac{\partial}{\partial \boldsymbol{\delta}} \sum_{i=1}^{n} K_{h_n}(Z_i - z) \left\{ \log(Y_i/w_n) \exp(\mathbf{X}_i^\top \boldsymbol{\theta} + \bar{\mathbf{Z}}_i^\top \boldsymbol{\delta}) - \mathbf{X}_i^\top \boldsymbol{\theta} - \bar{\mathbf{Z}}_i^\top \boldsymbol{\delta} \right\} I(Y_i > w_n) = 0. \tag{6}$$

For ease of notation, we write the solution to (6) as $\widehat{\boldsymbol{\delta}}_n = (\widehat{\delta}_{0,n}, \widehat{\delta}_{1,n})^\top$ with $\widehat{\delta}_{0,n} = \widehat{\eta}_n(\boldsymbol{\theta}, z)$ and $\widehat{\delta}_{1,n} = \widehat{\eta}_n^{(1)}(\boldsymbol{\theta}, z)$, and further estimate $\boldsymbol{\theta}$ by solving

$$\frac{\partial}{\partial \boldsymbol{\theta}} \sum_{i=1}^{n} \left\{ \log(Y_i/w_n) \exp(\mathbf{X}_i^\top \boldsymbol{\theta} + \widehat{\eta}_n(\boldsymbol{\theta}, Z_i)) - (\mathbf{X}_i^\top \boldsymbol{\theta} + \widehat{\eta}_n(\boldsymbol{\theta}, Z_i) \right\} I(Y_i > w_n) = 0. \tag{7}$$

Denote the corresponding estimator as $\widehat{\boldsymbol{\theta}}_n$. We then iterate between solving (6) and (7) until the estimates converge in the sense that $\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\| < \epsilon_1$ and $\|\widehat{\eta}_n(z) - \eta(z)\|_\infty < \epsilon_2$, where ϵ_1 and ϵ_2 are very small constants (e.g., 10^{-3}).

Note that the Hessian matrix of the objective function in (5) is $\sum_{i=1}^{n} K_{h_n}(Z_i - z)\log(Y_i/w_n)\exp(\mathbf{X}_i^\top\boldsymbol{\theta}+\bar{\mathbf{Z}}_i^\top\boldsymbol{\delta})(\mathbf{X}_i^\top,\bar{\mathbf{Z}}_i^\top)^\top(\mathbf{X}_i^\top,\bar{\mathbf{Z}}_i^\top)I(Y_i > w_n)$ which is positive definite for $\boldsymbol{\theta}$ and $\boldsymbol{\delta}$. This indicates the objective function is strictly convex, i.e., the local minimizer of (5) is its global minimizer. Thus, it is sufficient to construct the estimates by solving the likelihood equations.

Now we discuss the extension of the univariate case to multivariate ones by denoting the nonparametric component as $\eta_j(z), j = 1, \ldots, q$. The likelihood can be similarly derived as

$$\mathcal{L}_n(\boldsymbol{\theta}, \eta) = \sum_{i=1}^{n} \left\{ \log(Y_i/w_n) \exp\left[\mathbf{X}_i^\top\boldsymbol{\theta} + \sum_{j=1}^{q} \eta_j(Z_{ij})\right] - \mathbf{X}_i^\top\boldsymbol{\theta} - \sum_{j=1}^{q} \eta_j(Z_{ij}) \right\} I(Y_i > w_n). \tag{8}$$

An application of local smoothing on $\eta_j(Z_{ij})$ leads to

$$\sum_{i=1}^{n}\prod_{j=1}^{q} K_{h_n}(Z_{ij} - z_j) \left\{ \log(Y_i/w_n) \exp\left[\mathbf{X}_i^\top\boldsymbol{\theta} + \sum_{j=1}^{q} \bar{\mathbf{Z}}_{ij}^\top\boldsymbol{\delta}_j\right] - \mathbf{X}_i^\top\boldsymbol{\theta} - \sum_{j=1}^{q} \bar{\mathbf{Z}}_{ij}^\top\boldsymbol{\delta}_j \right\} I(Y_i > w_n),$$

where $\boldsymbol{\delta}_j = (\eta_j(z_j), \eta_j^{(1)}(z_j))^\top$ and the other notations are similarly defined as before. Estimating multiple nonparametric functions involves estimating η_j's, each of which depends on its individual tuning parameter, and can thus be computationally challenging. To overcome this, we can use the spline-backfitted local linear estimation approach in Liu and Yang (2010). Specifically, we can estimate the smooth functions via spline approximation first and re-estimate each of the additive functions using the profile local linear approach as described below.

Step 1. Using B-spline expansion of $\eta_j(Z_{ij})$ as $\eta_j(Z_{ij}) = \boldsymbol{\zeta}_j^\top\mathbf{B}(Z_{ij})$, we write (8) as

$$\sum_{i=1}^{n} \left\{ \log(Y_i/w_n) \exp\left[\mathbf{X}_i^\top\boldsymbol{\theta} + \sum_{j=1}^{q} \boldsymbol{\zeta}_j^\top\mathbf{B}(Z_{ij})\right] - \mathbf{X}_i^\top\boldsymbol{\theta} - \sum_{j=1}^{q} \boldsymbol{\zeta}_j^\top\mathbf{B}(Z_{ij})) \right\} I(Y_i > w_n),$$

then take its first derivative with respect to $(\boldsymbol{\theta}, \boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_q)$, leading to

$$\sum_{i=1}^{n} \left\{ \log(Y_i/w_n) \exp\left[\mathbf{X}_i^\top\boldsymbol{\theta} + \sum_{j=1}^{q} \boldsymbol{\zeta}_j^\top\mathbf{B}(Z_{ij})\right] - 1 \right\} [\mathbf{X}_i^\top, \mathbf{B}^\top(Z_i)]^\top I(Y_i > w_n) = 0,$$

where $\mathbf{B}(Z_i) = [\mathbf{B}^\top(Z_{i1}), \ldots, \mathbf{B}^\top(Z_{iq})]^\top$ and $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_q)^\top$. Then, we can derive the spline estimators $(\tilde{\boldsymbol{\theta}}, \tilde{\boldsymbol{\zeta}})$ by solving this equation.

Step 2. Integrating the estimator in Step 1 and local linear approximation of $\eta_j(Z_{ij})$, we obtain

$$\sum_{i=1}^{n} K_{h_n}(Z_{ij} - z_j)\{\log(Y_i/w_n)\exp[\mathbf{X}_i^\top \tilde{\boldsymbol{\theta}} + \sum_{j' \neq j}^{q} \tilde{\boldsymbol{\zeta}}_{j'}^\top \mathbf{B}(Z_{ij'}) + \bar{\mathbf{Z}}_{ij}^\top \boldsymbol{\delta}_j] \\ - \mathbf{X}_i^\top \tilde{\boldsymbol{\theta}} - \sum_{j' \neq j}^{q} \tilde{\boldsymbol{\zeta}}_{j'}^\top \mathbf{B}(Z_{ij'}) - \bar{\mathbf{Z}}_{ij}^\top \boldsymbol{\delta}_j\} I(Y_i > w_n).$$

Take its first derivative with respect to $\boldsymbol{\delta}_j$ leading to the following estimating equation

$$\sum_{i=1}^{n} K_{h_n}(Z_{ij} - z_j)\{\log(Y_i/w_n)\exp[\mathbf{X}_i^\top \tilde{\boldsymbol{\theta}} + \sum_{j' \neq j}^{q} \tilde{\boldsymbol{\zeta}}_{j'}^\top \mathbf{B}(Z_{ij'}) + \bar{\mathbf{Z}}_{ij}^\top \boldsymbol{\delta}_j] - 1\}\bar{\mathbf{Z}}_{ij} I(Y_i > w_n) = 0.$$

Then the second stage estimators $\widehat{\eta}_j(z_j)$ follows and the estimator $\widehat{\boldsymbol{\theta}}$ can be further derived similarly. We repeat Step1 and Step 2 iteratively until convergence.

Since our focus is to estimate the parameters in the STIR model and understand the properties of their estimates, for the remaining part of the paper we focus on a univariate function η. Nevertheless, we have implemented the estimation procedure elaborated above and will apply it to the data analysis in Section 5.2.

3 Asymptotic theory

In this section, we investigate the asymptotic properties of the proposed estimators under the assumptions below, with the technical proofs presented in the Appendix.

(C1) The variables $(\mathbf{X}_i, Z_i)$ are independent and identically distributed on the compact set $\mathcal{X} \otimes \mathcal{Z} \subset \mathbb{R}^p \otimes \mathbb{R}^1$, while the marginal density $f_Z(z)$ of Z is uniformly bounded away from 0 and infinity, i.e., there exist constants $\underline{c}$ and $\overline{c}$ such that

$$0 < \underline{c} \leq \inf_{z \in \mathcal{Z}} f_Z(z) \leq \sup_{z \in \mathcal{Z}} f_Z(z) \leq \overline{c}.$$

(C2) The matrix $\mathrm{E}(\mathbf{X}_i \mathbf{X}_i^\top)$ is positive definite for each $i = 1, \ldots, n$.

(C3) The kernel $K(z)$ is a symmetric and continuous probability density function satisfying $\int_{\mathcal{Z}} K(z)\,dz = 1$, $\int_{\mathcal{Z}} z^2 K(z)\,dz < \infty$, and $\mid z \mid K(z) \to 0$ as $\mid z \mid \to \infty$.

(C4) The bandwidth h_n satisfies $h_n \to 0$, $n_0 h_n^2 \to 0$, $n/n_0^2 h_n \to 0$, $nh_n^2/n_0 \to 0$, and $n_0 h_n / \log(n_0) \to \infty$ as the sample size approaches infinity.

(C5) Let τ, b, and ϱ be real constants satisfying $0 < b \leq 1$ and $\varrho > 0$, while $\eta(z)$ is τ-order continuously differentiable function for $\tau \geq 2$ such that

$$\mid \eta^{(\tau-2)}(z_1) - \eta^{(\tau-2)}(z_2) \mid \leq \varrho \mid z_1 - z_2 \mid^b,$$

where $\eta^{(\tau-2)}(z)$ is the $(\tau-2)$th derivative of $\eta(z)$.

(C6) The term $o(y^{-\beta(\mathbf{x},z)})$ in the slowly varying function $L(y;\mathbf{x},z)$ satisfies that when $y \to \infty$,

$$\sup_{\mathbf{x}\in\mathcal{X}, z\in\mathcal{Z}} \{y^{\beta(\mathbf{x},z)} o(y^{-\beta(\mathbf{x},z)})\} \to 0.$$

(C7) The matrices $\boldsymbol{\Xi}_{kk}(z), k = 1,2$ and $\boldsymbol{\Pi}$ defined in Section 3 are nonsingular, while $\boldsymbol{\Lambda}_{11}(z)$ and $\mathbf{V}$ are positive definite.

(C8) The consistent estimator $\widehat{\boldsymbol{\delta}}_n(\boldsymbol{\theta}, z)$ has bounded second order continuous derivative uniformly and satisfies $\|\widehat{\boldsymbol{\delta}}_n^{(j)}(\boldsymbol{\theta}, z) - \boldsymbol{\delta}^{(j)}(\boldsymbol{\theta}, z)\|_\infty \to 0$ in probability for $j = 0, 1, 2$, where $\|\eta(z)\|_\infty = \sup_{z\in\mathcal{Z}} \mid \eta(z) \mid$.

Remark 1. Assumptions (C1) and (C2) are commonly used in regression models, while Conditions (C3) and (C4) on the kernel function and the bandwidth are typical in non/semi-parametric estimation when local polynomial smoothing is employed. The Lipschitz continuous condition in (C5) guarantees the smoothness of $\eta(z)$ in estimation and the asymptotic theory. Following condition (C2) in Wang and Tsai (2009), we use (C6) to regularize the extreme behavior of the slowly varying function. For simplicity, Conditions (C7) and (C8) are needed in establishing asymptotic properties of the parametric and nonparametric estimators.

For ease of notation, we write

$$\mu_k = \int_{\mathcal{Z}} z^k K(z)dz,\ \nu_k = \int_{\mathcal{Z}} z^k K^2(z)dz \text{ for } k = 0,1,2,3, \quad \text{and}$$

$$\Re(\mathbf{X}_1, Z_1) = c_0(\mathbf{X}_1, Z_1)w_n^{-\alpha(\mathbf{X}_1,Z_1)} + \frac{c_1(\mathbf{X}_1, Z_1)\alpha(\mathbf{X}_1, Z_1)}{\alpha(\mathbf{X}_1, Z_1) + \beta(\mathbf{X}_1, Z_1)} w_n^{-\alpha(\mathbf{X}_1,Z_1)-\beta(\mathbf{X}_1,Z_1)},$$

and assume that the following expressions are convergent in probability when the sample size $n \to \infty$, i.e.,

$$\frac{n}{n_0}\mu_{2(k-1)}\mathrm{E}\left\{\Re(\mathbf{X}_1, Z_1) \mid Z_1 = z\right\} f_Z(z) \to_p \boldsymbol{\Xi}_{kk}(z)\ ,\ k = 1, 2; \tag{9}$$

$$\sqrt{\frac{n^2 h_n}{n_0}} \left\{\frac{h_n^2}{2}\eta^{(2)}(z)\mu_2 - 1\right\} \mathrm{E}\left[\Re(\mathbf{X}_1, Z_1) + c_0(\mathbf{X}_1, Z_1)w_n^{-\alpha(\mathbf{X}_1,Z_1)}|Z_1 = z\right] f_Z(z) \to_p \boldsymbol{\Sigma}_{11}(z); \tag{10}$$

$$\begin{aligned} &\frac{n}{n_0}\mathrm{E}\left[c_0(\mathbf{X}_1, Z_1)w_n^{-\alpha(\mathbf{X}_1,Z_1)} + c_1(\mathbf{X}_1, Z_1)\frac{\alpha^2(\mathbf{X}_1, Z_1) + \beta^2(\mathbf{X}_1, Z_1)}{\{\alpha(\mathbf{X}_1, Z_1) + \beta(\mathbf{X}_1, Z_1)\}^2} w_n^{-\alpha(\mathbf{X}_1,Z_1)} \right. \\ &\left. \cdot w_n^{\beta(\mathbf{X}_1,Z_1)}|Z_1 = z\right] f_Z(z)\nu_0 \to_p \boldsymbol{\Lambda}_{11}(z), \end{aligned} \tag{11}$$

where $f_Z(z)$ is the marginal probability density function of Z. In the text followed, let $\boldsymbol{\theta}_0$ and $\eta_0(z)$ be the true values of $\boldsymbol{\theta}$ and $\eta(z)$ respectively.

Theorem 1. *Suppose that Conditions (C1)–(C8) are satisfied. For random variable Z over a compact set $\mathcal{Z}$ and $\sqrt{n_0}$-consistent solution $\widehat{\boldsymbol{\theta}}_n$ to* (6) *satisfying $\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0 = O_p(1/\sqrt{n_0})$, we have that, for given $Z = z$,*

$$\sqrt{n_0 h_n}\left\{\widehat{\eta}_n(\widehat{\boldsymbol{\theta}}_n, z) - \eta_0(z) - \boldsymbol{\Xi}_{11}^{-1}(z)\boldsymbol{\Sigma}_{11}(z)\right\} \to_d \mathrm{N}(0, \varrho)$$

converges in distribution as $n_0 \to \infty$, where $\varrho = \boldsymbol{\Xi}_{11}^{-1}(z)\boldsymbol{\Lambda}_{11}(z)\boldsymbol{\Xi}_{11}^{-1}(z)$ with $\boldsymbol{\Xi}_{11}(z)$, $\boldsymbol{\Sigma}_{11}(z)$ and $\boldsymbol{\Lambda}_{11}(z)$ defined in (9), (10) *and* (11) *respectively.*

Theorem 1 presents the asymptotic normality of the estimator of the nonparametric function $\eta(z)$ once a $\sqrt{n_0}$-consistent estimator $\widehat{\boldsymbol{\theta}}_n$ of $\boldsymbol{\theta}$ can be derived, which is available in the following proposition. To this end, we assume that

$$\frac{n}{n_0}\mathrm{E}\left[\{\mathbf{X}_1 + \bar{\mathbf{Z}}_1^{\top}\widehat{\boldsymbol{\delta}}_n^{(1)}(\boldsymbol{\theta}_0, Z_1)\}\{\mathbf{X}_1 + \bar{\mathbf{Z}}_1^{\top}\widehat{\boldsymbol{\delta}}_n^{(1)}(\boldsymbol{\theta}_0, Z_1)\}^{\top}\Re(\mathbf{X}_1, Z_1)\right] \to \boldsymbol{\Pi}_1, \tag{12}$$

and

$$\frac{n}{n_0}\mathrm{E}\left[\bar{\mathbf{Z}}_1^\top \widehat{\boldsymbol{\delta}}_n^{(2)}(\boldsymbol{\theta}_0, Z_1)\left\{\Re(\mathbf{X}_1, Z_1) - c_0(\mathbf{X}_1, Z_1) w_n^{-\alpha(\mathbf{X}_1, Z_1)}\right\}\right] \to \boldsymbol{\Pi}_2. \tag{13}$$

Theorem 2. *Suppose that Conditions (C1)–(C8) are satisfied and $nh_n^4 \to 0$, then the estimator $\widehat{\boldsymbol{\theta}}_n$ in Section 2 converges in distribution as $n_0 \to \infty$, i.e.,*

$$\sqrt{n_0}(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) \to_d \mathrm{N}(\boldsymbol{\zeta}, \boldsymbol{\Pi}^{-1}\mathbf{V}\boldsymbol{\Pi}^{-1}),$$

where $\boldsymbol{\Pi} = \boldsymbol{\Pi}_1 + \boldsymbol{\Pi}_2$ are defined in (12) *and* (13)*, and*

$$\mathbf{V} = \mathrm{E}\left[\{\mathbf{X}_1 + \eta^{(1)}(\boldsymbol{\theta}_0, Z_1)\}\{\mathbf{X}_1 + \eta_n^{(1)}(\boldsymbol{\theta}_0, Z_1)\}^\top \mid Y_1 > w_n\right].$$

Moreover, the bias term $\boldsymbol{\zeta}$ has the form $\boldsymbol{\zeta} = \boldsymbol{\zeta}_1 + \boldsymbol{\zeta}_2$ in which

$$\begin{aligned}
\boldsymbol{\zeta}_1 &= \lim_{n_0\to\infty} -\frac{n}{\sqrt{n_0}}\mathrm{E}\left[\{\mathbf{X}_1 + \widehat{\eta}_n^{(1)}(\boldsymbol{\theta}_0, Z_1)\}\frac{c_1(\mathbf{X}_1, Z_1)\beta(\mathbf{X}_1, Z_1)}{\alpha(\mathbf{X}_1, Z_1) + \beta(\mathbf{X}_1, Z_1)} w_n^{-\alpha(\mathbf{X}_1, Z_1) - \beta(\mathbf{X}_1, Z_1)}\right], \\
\boldsymbol{\zeta}_2 &= \lim_{n_0\to\infty} -\frac{n}{\sqrt{n_0}}\frac{h_n^2}{2}\mathrm{E}\left[\eta_n^{(2)}(\boldsymbol{\theta}_0, Z_1)\left\{c_0(\mathbf{X}_1, Z_1) w_n^{-\alpha(\mathbf{X}_1, Z_1)} + \frac{c_1(\mathbf{X}_1, Z_1)\alpha(\mathbf{X}_1, Z_1)}{\alpha(\mathbf{X}_1, Z_1) + \beta(\mathbf{X}_1, Z_1)}\right.\right. \\
&\quad \left.\left.\cdot w_n^{-\alpha(\mathbf{X}_1, Z_1) - \beta(\mathbf{X}_1, Z_1)}\right\}\{\mathbf{X}_1 + \widehat{\eta}_n^{(1)}(\boldsymbol{\theta}_0, Z_1)\}\right].
\end{aligned}$$

Theorem 2 shows the asymptotical normality of $\widehat{\boldsymbol{\theta}}_n$ with a non-ignorable bias term similarly to that in Wang and Tsai (2009). This bias can be reduced if $\boldsymbol{\zeta}_2$ can be removed. This happens if $nh_n^2/\sqrt{n_0} \to 0$ as $n_0 \to \infty$ which is satisfied by undersmoothing when $h_n = o((n_0/n^2)^{1/4})$ and particularly $h_n = o(n^{-1/4})$ when the number of selected observations n_0 has same order to total sample size n. We remark that the matrices $\boldsymbol{\Pi}$ and $\mathbf{V}$ can be consistently estimated by their sample analogs.

4 Implementation

In this section, we discuss how to choose the tuning parameters and provide the estimation procedure.

4.1 Tuning parameter selection

The estimating procedure relies on two tuning parameters including the threshold w_n and the bandwidth h_n, the latter of which determining the smoothness of the fitted curve $\eta(\cdot)$ is affected by the tuning parameter w_n controlling the effective sample size n_0.

We discuss the choice of w_n first. We note that the distribution of $\exp(\mathbf{X}_i^\top \boldsymbol{\theta} + \eta(Z_i)) \log(Y_i/w_n)$ conditional on $Y_i > w_n$ is approximately standard exponential, and the negative exponent of $U_i = \exp(-\exp(\mathbf{X}_i^\top \boldsymbol{\theta} + \eta(Z_i)) \log(Y_i/w_n))$ has an approximate uniform distribution over $[0, 1]$. Then, using the distance function proposed in Wang and Tsai (2009) that measures the difference between the theoretical uniform distribution and the empirical one, an appropriate w_n can be determined by minimizing the following distance

$$\widehat{D}(w_n, h_n) = n_0^{-1} \sum_{i=1}^{n} \{\widehat{U}_i(h_n) - \widehat{F}_n(U_i)\}^2 I(Y_i > w_n), \tag{14}$$

where $\widehat{F}_n(\cdot)$ is the empirical distribution of U_i and

$$\widehat{U}_i(h_n) = \exp[-\exp\{\mathbf{X}_i^\top \widehat{\boldsymbol{\theta}}_n + \widehat{\eta}_n(Z_i)\} \log(Y_i/w_n)].$$

We note that $\widehat{D}(w_n, h_n)$ is a function of h_n as $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\eta}_n(Z_i)$ depend on h_n. In reality, we scan all the values of Y_i and choose the optimal w_n that minimizes $\widehat{D}(w_n, h_n)$ for a fixed h_n.

For the choice of the bandwidth h_n, one may resort to cross validation (Rudemo, 1982; Stone, 1984; Bowman, 1984) or the plug-in principle in Silverman (1986). In our numerical experiments, it is found that the local linear estimator of $\eta(z)$ is not very sensitive to the choice of h_n once its order is suitably determined based on sample size n_0. Thus we simply adopt the plug-in bandwidth by choosing an optimal h_n in the set $\{h_n : h_n = c * h_{rot}, c > 0\}$ with $h_{rot} = 1.06 \widehat{\sigma}_z n_0^{-1/5}$, where $\widehat{\sigma}_z$ is the standard error of z based on the used n_0 observations. It is worth mentioning that undersmoothing is often employed to obtain $\sqrt{n}$-consistent estimator of parametric components in semiparametric regression; see Wong et al. (2008), Li et al. (2011) and Sun et al. (2012) among many others. To this end in our context, we use the plug-in method above by taking $c = c_0 * n_0^{-\epsilon}$ with $\epsilon = 1/15$ and constant $c_0 > 0$ so that the undersmoothing condition is satisfied as $h_n = o(n^{-1/4})$.

4.2 Computation

For the estimation of $\boldsymbol{\theta}$ and $\eta(z)$, we summarize the computing procedure in four steps.

- Step 1. Initial estimator $\widehat{\boldsymbol{\theta}}_{n,0}$. We use the method proposed in Wang and Tsai (2009) by assuming that $\eta(z) = a + bz$ has a linear structure.

- Step 2. Update $\eta(z)$. With the initial estimator $\widehat{\boldsymbol{\theta}}_{n,0}$, we apply local linear smoothing to solve (6). Denote the solution as $\widehat{\boldsymbol{\delta}}_n(\widehat{\boldsymbol{\theta}}_{n,0}, z)$ with $z = Z_i$ for $i = 1, \ldots, n$.

- Step 3. Update $\boldsymbol{\theta}_{n,0}$. We replace $\boldsymbol{\delta}$ by $\widehat{\boldsymbol{\delta}}(\widehat{\boldsymbol{\theta}}_{n,0}, z)$ and solve equation (7) using Fisher scoring algorithm to obtain

$$\boldsymbol{\zeta}_n + \boldsymbol{\Pi}_n \sqrt{n_0}(\widehat{\boldsymbol{\theta}}_{n,1} - \boldsymbol{\theta}_0) + o_p(1) = 0,$$

  where $\boldsymbol{\zeta}_n$ and $\boldsymbol{\Pi}_n = \boldsymbol{\Pi}_{1n} + \boldsymbol{\Pi}_{2n}$ are matrices defined in the Appendix.

- Step 4. Repeat Step 2 and Step 3 until $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\eta}_n(z)$ converge.

5 Numerical experiments

In this section, simulations and two real data analyses are conducted to illustrate the finite sample performance of proposed method.

5.1 Simulations

Example 1. In this example, 200 datasets were generated from the following model

$$1 - F(y; \mathbf{x}, z) = \frac{(1+\gamma)y^{-\alpha(\mathbf{x},z)}}{1+\gamma y^{-\alpha(\mathbf{x},z)}} = y^{-\alpha(\mathbf{x},z)} \left\{(1+\gamma) - \gamma(1+\gamma)y^{-\alpha(\mathbf{x},z)} + o(y^{-\alpha(\mathbf{x},z)})\right\},$$

where $\log(\alpha(\mathbf{x}, z)) = \mathbf{x}^\top \boldsymbol{\theta} + \eta(z)$, the parameter γ in the slowly varying function is taken as $\gamma = 0.15, 0.3$ or 0.6, and the sample size is $n = 200, 400$ or 800.

For the parametric component, $\mathbf{X}_i, i = 1, \ldots, n$ are cross-sectionally independent and identically distributed variables from $\mathbf{X}_i \sim 0.5\mathrm{N}(\mathbf{0}, \boldsymbol{\Omega})$, where the (k, j)th entry of $\boldsymbol{\Omega}$ is

$\omega_{k,j} = 0.5^{|k-j|}$ for $k, j = 1, \ldots, 4$. The coefficient vector is set as $\boldsymbol{\theta}_0 = (0.6, -0.4, 0.3, -0.5)^\top$. For the nonparametric part, we consider either of the following two smooth functions

$$\eta_1(z) = \sin(2z) - 0.2\exp(-16z^2) \quad \text{or} \quad \eta_2(z) = \exp(-5z - 1) + z^2,$$

in which z is generated from a truncated standard normal distribution over $[-2, 2]$ for $\eta_1(z)$ and from an uniform distribution over $[0, 1]$ for $\eta_2(z)$, respectively. The setting is adopted such that the generated data is similar to the real data analyzed in Section 5.2. In particular, the left panel in Figure 1 presents the histogram of simulated observations of the response variable when the function $\eta_1(z)$ is used, in which many extreme values are observed, where the right panel of Figure 1 displays the empirical distribution of the ROE data analyzed in Section 5.2. To determine the sample fraction used, we provide 50 $w_n's$ with the corresponding sample fractions equally distributed over $[1/3, 1]$ and select an optimal one using the criterion in (14) when h_n is fixed.

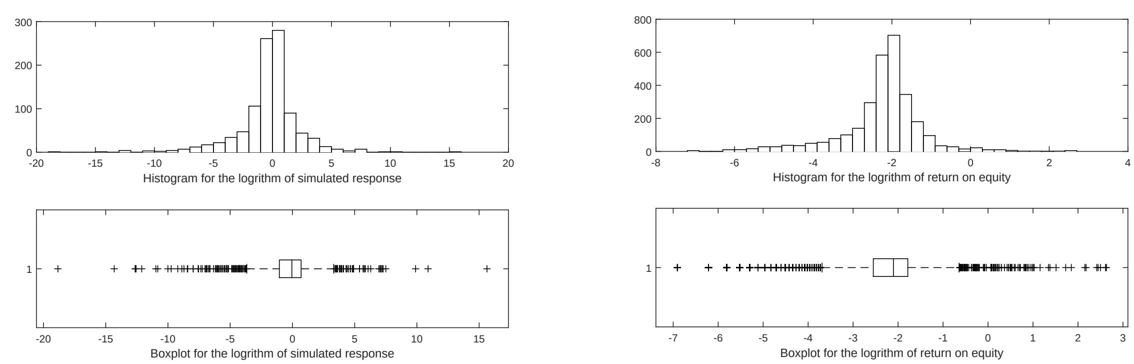

Figure 1: Histograms and box plots for the logarithm of absolute response $\mid y \mid$ in simulations (left panel) and that of the logarithm of absolute return on equity | ROE | (right panel).

The goodness of fit for $\widehat{\boldsymbol{\theta}}_n$ and $\widehat{\eta}_n(z)$ is measured by mean squared errors defined for parametric and nonparametric parts respectively as

$$\begin{aligned}
\mathrm{MSE}_p &= \frac{1}{n_0}\|\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\|^2, \\
\mathrm{MSE}_{np} &= \frac{1}{n_0}\int_a^b \{\widehat{\eta}_n(z) - \eta(z)\}^2\, dz,
\end{aligned}$$

where MSE_{np} is evaluated at 101 equally spaced grids on $[a, b] \in \{[-1.8, 1.8], [0, 1]\}$, and the

Table 1: Simulation results: γ is the parameter in the slowly varying function; ω_n is sample fraction determined by the effective sample size n_0 as n_0/n; MSE_p and MSE_{np} are empirical average of the mean square error for parametric and nonparametric estimates; STD_p and STD_{np} are the corresponding empirical standard deviations respectively.

γ	n	ω_n	MSE_p	STD_p	MSE_{np}	STD_{np}
		$\eta_1(z) = \sin(2z) - 0.2\exp(-16z^2)$				
0.15	200	0.73	0.04	0.09	0.31	0.06
	400	0.67	0.02	0.08	0.30	0.05
	800	0.60	0.01	0.06	0.28	0.03
0.30	200	0.76	0.04	0.12	0.32	0.06
	400	0.75	0.04	0.10	0.31	0.05
	800	0.70	0.01	0.07	0.29	0.03
0.60	200	0.82	0.03	0.13	0.36	0.07
	400	0.80	0.02	0.09	0.35	0.04
	800	0.79	0.01	0.06	0.33	0.03
		$\eta_2(z) = \exp(-5z - 1) + z^2$				
0.15	200	0.76	0.03	0.06	0.02	0.04
	400	0.71	0.02	0.02	0.01	0.02
	800	0.70	0.01	0.02	0.01	0.01
0.30	200	0.71	0.05	0.06	0.03	0.04
	400	0.66	0.03	0.04	0.01	0.01
	800	0.63	0.02	0.01	0.01	0.01
0.60	200	0.69	0.06	0.06	0.04	0.03
	400	0.62	0.03	0.03	0.02	0.02
	800	0.50	0.02	0.03	0.01	0.01

truncated normal kernel function $K(z) = \exp(-0.5 \times z^2)/\sqrt{2\pi} I(|z| < 15)$ is used in the local linear estimator $\widehat{\eta}_n(z)$. The bandwidth h_n is selected by minimizing $\widehat{D}(w_n, h_n)$ in (14) from a grid equally spaced on $[h_{\mathrm{rot}}/2, 2h_{\mathrm{rot}}]$ for a fixed w_n, where $h_{rot} = 1.06\widehat{\sigma}_z n_0^{-1/5}$ is the rule-of-thumb bandwidth proposed in Silverman (1986) and $\widehat{\sigma}_z$ is the estimated standard deviation of z and n_0 is the effective sample size accordingly. We summarize the semiparametric tail index regression results in Table 1 and Table 2.

In Table 1, we observe that the effective sample size in estimation increases at a a slower rate than the sample size. The mean squared errors MSE_p and MSE_{np} are decreasing simultaneously as n grows, while the standard deviations in STD_p and STD_{np} have similar performance. These results indicate the consistency of our STIR estimators. In Table 2, we report the biases and standard deviations of the estimated coefficients when either η_1 or

Table 2: Biases (bias) and standard deviations (std) of $\widehat{\boldsymbol{\theta}}_n$, where γ is the parameter in the slowly varying function.

	γ	0.15			0.3			0.6		
	n	200	400	800	200	400	800	200	400	800
		$\eta_1(z) = \sin(2z) - 0.2\exp(-16z^2)$								
$\widehat{\theta}_{1,n}$	bias	0.00	0.02	0.02	0.01	0.00	0.01	0.01	0.02	0.02
	std	0.22	0.15	0.10	0.22	0.13	0.11	0.20	0.14	0.10
$\widehat{\theta}_{2,n}$	bias	-0.01	0.00	0.01	-0.03	0.00	-0.02	0.00	-0.05	-0.02
	std	0.24	0.16	0.13	0.23	0.16	0.13	0.21	0.15	0.11
$\widehat{\theta}_{3,n}$	bias	0.02	-0.03	0.00	0.02	-0.01	0.02	0.01	0.03	0.02
	std	0.25	0.16	0.14	0.23	0.17	0.12	0.22	0.16	0.10
$\widehat{\theta}_{4,n}$	bias	-0.05	0.00	-0.02	-0.04	0.01	-0.03	-0.03	-0.01	-0.02
	std	0.23	0.16	0.11	0.22	0.13	0.11	0.21	0.13	0.10
		$\eta_2(z) = \exp(-5z - 1) + z^2$								
$\widehat{\theta}_{1,n}$	bias	0.02	0.00	-0.03	0.03	0.00	-0.02	0.05	0.03	-0.01
	std	0.19	0.16	0.11	0.19	0.16	0.11	0.23	0.18	0.13
$\widehat{\theta}_{2,n}$	bias	0.01	0.00	0.01	-0.04	0.03	0.00	-0.04	-0.03	0.00
	std	0.21	0.18	0.12	0.24	0.16	0.12	0.24	0.15	0.13
$\widehat{\theta}_{3,n}$	bias	-0.01	0.01	-0.01	0.02	-0.02	0.01	0.04	0.01	0.00
	std	0.21	0.19	0.11	0.24	0.16	0.14	0.26	0.19	0.14
$\widehat{\theta}_{4,n}$	bias	-0.02	-0.01	0.02	-0.01	0.01	0.00	-0.07	-0.02	0.01
	std	0.21	0.18	0.12	0.25	0.14	0.12	0.20	0.17	0.13

η_2 is used as the nonparametric function. We can see that both quantities decrease to zero as the sample size increases. Graphically, the left panel in Figure 2 displays the empirical distribution of $\widehat{\boldsymbol{\beta}}_n$ based on 200 realizations, implying that $\widehat{\boldsymbol{\beta}}_n$ follows a normal distribution, while the right panel shows the fitted curves and the corresponding 95% pointwise confidence bands for $\eta_1(z)$ or $\eta_2(z)$, indicating that the nonlinear functions are fitted well.

One interesting questions arises as to whether the minimum of the distance function defined in (14) exists. Towards answering this, we plot $\widehat{D}(w_n, h_n)$ as a function of w_n and h_n using different sample sizes and different values for γ in the slowly varying function. As can be seen from Figure 3, we can see a clear optimal combination of w_n and h_n that minimizes the distance function.

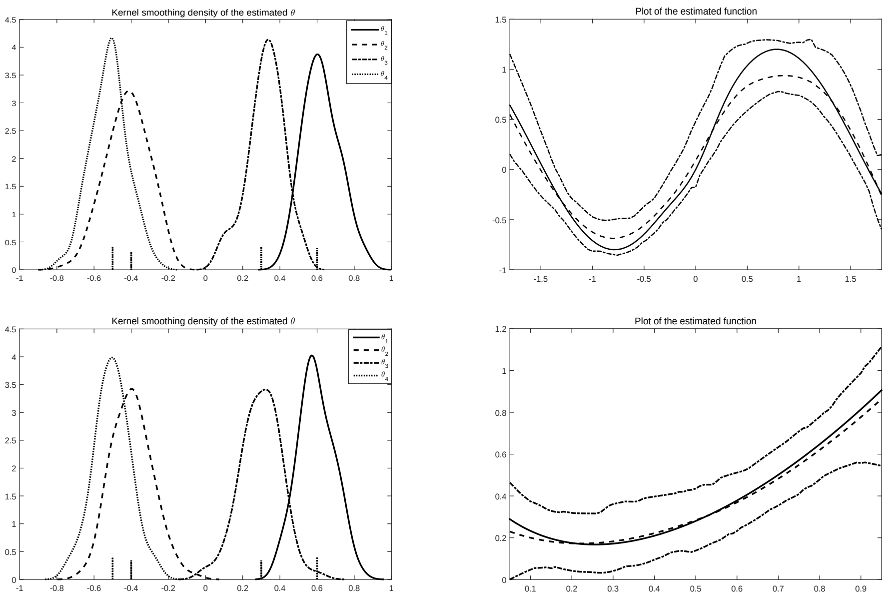

Figure 2: Empirical distribution of the estimator $\widehat{\boldsymbol{\theta}}_n$ with either $\eta_1(z)$ or $\eta_2(z)$ in the simulated model are displayed in the left panel; Fitted curves for smooth functions (dashed line) and their 95% pointwise confidence bands (dashed dotted line) are displayed in the right panel in which the upper solid curve is the true $\eta_1(z)$ and the lower solid curve is the true $\eta_2(z)$.

5.2 Real data analysis

Example 2. (ROE data)

Return on equity measures the efficiency of a firm at generating profits from equity and shows how well a company generates earning growth. Considered as desirable when above 15%, it is an important metric for firms and is also one of the most important indices in investment decisions which is commonly affected by the following five economic variables. The first is profit margin ratio (PM), which can be seasonal, showing how efficiently a business operates. The second is asset turnover ratio (ATR) measuring how efficient a company is at using its assets to generate profits. More specifically, companies with low profit margins tend to have a high asset turnover, while those with high profit margins have a low asset turnover. Financial leverage (LEV) is the third variable referring to the amount of debt a company has for financing its operations, which perhaps has nonlinear impact on a firm's development since LEV may generate profit and undertake finical risk simultaneously. That is, the influence of LEV on ROE may fluctuate and be accompanied by uncertainty. Moreover, a large LEV value carries a risk of bankruptcy. The fourth variable

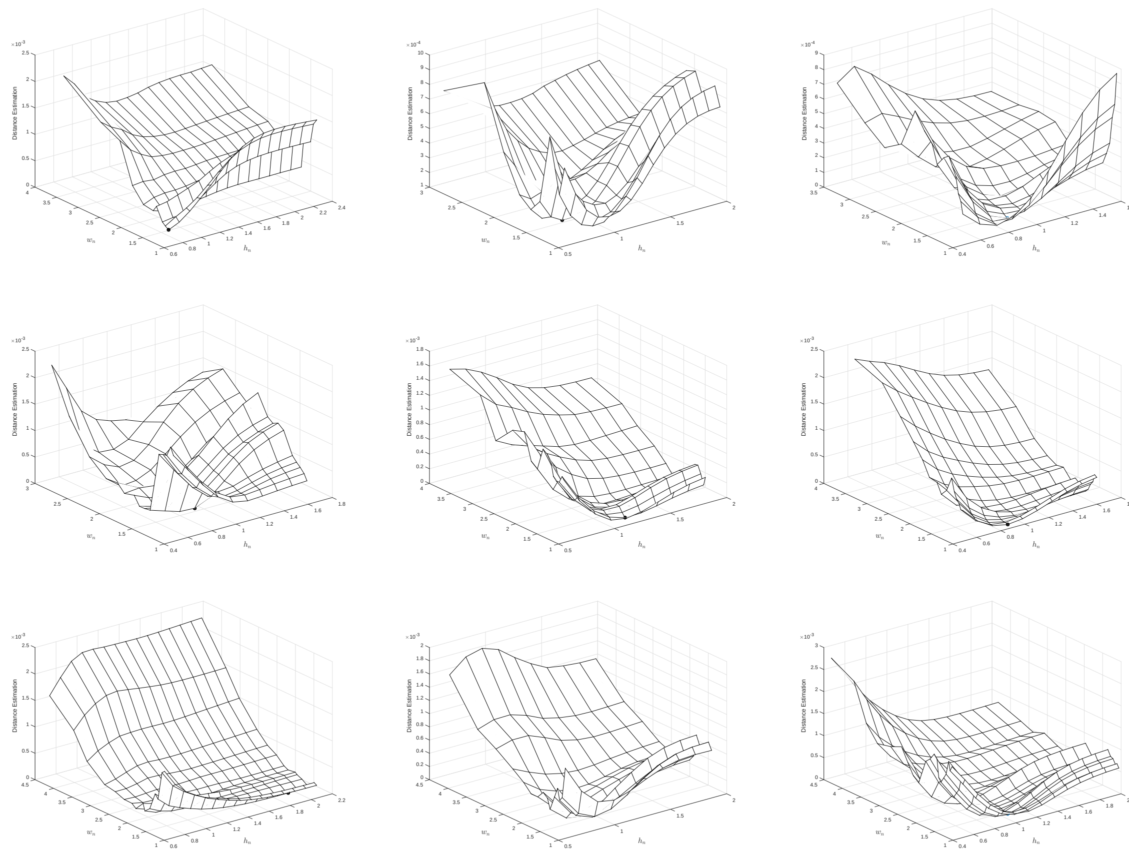

Figure 3: The distance function $\widehat{D}(w_n, h_n)$ in equation (14) when the parameter γ in the slowly varying function is taken to be $0.15, 0.3$ or 0.6 (from left to right), and the sample size n is set to be $n = 200, 400$, or 800 (from top to bottom).

is total asset (ASSET) as an economic resource from which the future economic profit could be guaranteed. The final variable is historical sales growth rate (SGR), one of the simplest approaches for estimating future growth. Thus, any changes in these economic indices make it difficult to predict the future return on equity. Motivated by these considerations, we use our semiparametric model to study the tail behavior of the return a firm on equity by taking the effect of LEV as nonlinear in an unknown functional form, while modelling the other variables linearly suggested by the monotone trend in preliminary analysis.

In this application, we use the data set in Wang and Tsai (2009) and analyze 2932 observations collected from 1997 to 2000 after standardization. Because we are mainly interested in the extreme values regardless of whether they are positive or negative, we take the absolute value of ROE as the response variable similarly to Wang and Tsai (2009). The histogram and box-plot are shown in the right panel of Figure 1. We remark that the raw data shows that many firms were booming while many other firms were very close to bankruptcy. A natural question arises then. Why did some companies have greater return volatility than the others though they were in same market circumstances?

Using our methodology, the effective sample size is found to be $n_0 = 585$ determined by the selected threshold $w_n = 0.185$ through minimizing the distance criterion in (14), giving a sample fraction $n_0/n = 19.95\%$. That is, those companies with ROE larger than 0.185 or less than -0.185 are the objects in our tail analysis. Although more detailed information of such companies is inaccessible, we argue that the companies with return on equity less than -0.185 faced huge operating risks in fierce market competition and, as a result, tended to fall during the Asian financial crisis from 1997 to 1998. On the other hand, those companies with return on equity larger than 0.185 also endured risk of bankruptcy when facing serious financial crisis. Therefore, the selected companies in our tail analysis can be considered extreme in some sense.

For the parametric component, the tail index estimates are reported in the upper panel of Table 3. We observe that both ASSET and ATR have significant positive effects and conclude that larger ASSET is helpful to the development of a company and that the enterprise with higher ATR can use its asset more effectively for generating profits. These conclusions are consistent with the definition of ROE and ATR. More specifically, ROE is positively proportional to company's net income and ATR is similarly related to net sales. More importantly, net income is part of new sales obviously indicating that the rise of ATR promotes the growth of ROE. In addition, although the equity ASSET is different from net income, it is very likely one of the factors leading to the increase of total asset and perhaps vise verse. Moreover, the net income consists of PM and salaries of employees among others. Thus, it is to be expected that PM also has positive impact on return on equity. Finally, for SGR, many companies grew at a relatively high speed during a extended period of time. These companies could avoid the shortage of funds by increasing equity capital and investment. Issuing new shares and streamlining internal finance are popular approaches commonly used. Thus, it is natural for SGR and ROE to have tendency to change in the same direction. Although PM and SGR have insignificant effects on ROE in our analysis possibly due to collinearity among these economic variables, their performances seem highly consistent with the economic principles.

Table 3: Estimates of the parameters in ROE data analysis and the corresponding empirical standard deviation (std), t statistics value (t-stat) and p-value.

covariate	estimate	std	t-stat.	p-value
	$\log(\alpha(\mathbf{x}, z)) = \mathbf{x}^\top \boldsymbol{\theta} + \eta(z)$			
ASSET	0.1127	0.0340	3.3205	0.0009
PM	0.0416	0.0215	1.7426	0.0814
SGR	0.0145	0.0220	0.5847	0.5587
ATR	0.0929	0.0295	3.1920	0.0014
	$\log(\alpha(\mathbf{x}, z)) = \mathbf{x}^\top \boldsymbol{\theta} + z\theta_{p+1}$			
ASSET	0.0940	0.0519	1.8302	0.0672
PM	0.0365	0.0167	2.1967	0.0280
SGR	0.0298	0.0287	0.9701	0.3320
ATR	0.2164	0.0303	7.1513	0.0000
LEV	-0.1270	0.0319	3.9012	0.0001
	$\log(\alpha(\mathbf{x}, z)) = \mathbf{x}^\top \boldsymbol{\theta} + z\theta_{p+1} + z^2\theta_{p+2} + z^3\theta_{p+3}$			
ASSET	0.0964	0.0479	2.0251	0.0429
PM	0.0350	0.0197	1.8264	0.0678
SGR	0.0279	0.0244	1.0035	0.3156
ATR	0.2016	0.0311	6.5064	0.0000
LEV	0.6585	0.4184	1.4735	0.1406
LEV^2	-1.3795	0.6665	1.9636	0.0496
LEV^3	0.8070	0.3564	2.1578	0.0309

For the nonparametric component, we display the fitted curve for LEV and the 95% pointwise bootstrap-based confidence bands in the left panel of Figure 4, in which the nonlinear effect of LEV on return on equity is quite clear. To interpret the fitted nonlinear trend, we see that LEV enables a company to control resources larger than their own equity capital. As a result, whether the company profits from additional debt or sustains economic loss, due to the imbalance between return on investment and the interest rate on borrowed capital, often results in great return volatility, particularly for those companies with high LEV. Motivated by the fitted LEV curve in Figure 4, we use a cubic polynomial to fit LEV and report the resulting estimates in the lower panel of Table 3. We see that LEV^3 has significant effect on return of equity. We also observe that the linear and the quadratic

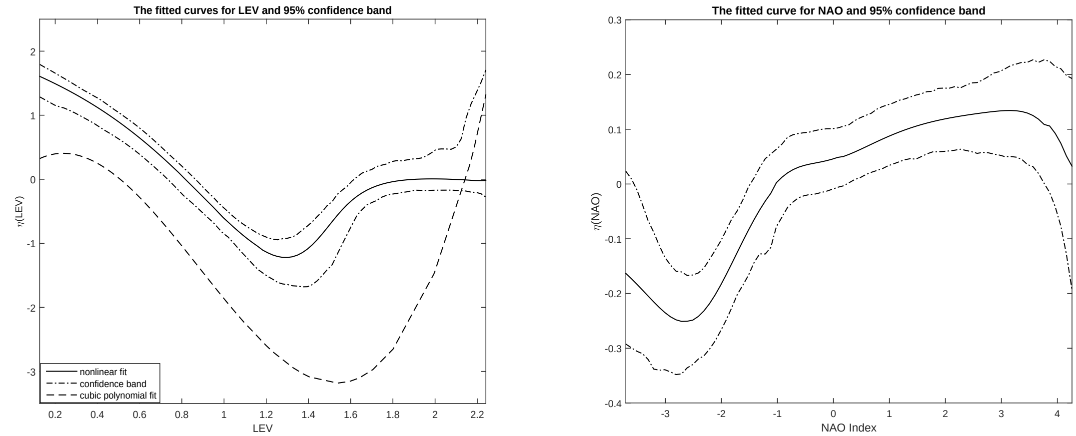

Figure 4: Fitted curves (solid line) for the estimated functions η(LEV) in left panel and η(NAO) in right panel; The 95% bootstrap confidence bands (dot-dashed line) and cubic polynomial fitted curve for η(LEV) in left panel (dashed line).

terms of LEV are almost insignificant possibly due to the collinearity among the three terms related to LEV. Although the cubic polynomial fit appears to be able to capture the overall trend of the nonlinear function η(LEV) as shown in the left panel of Figure 4, it can not model the fine details and provide a fit as flexible as our STIR model. Finally, for comparison purposes, we also report the results in the middle panel of Table 3 by including LEV in the parametric part of the model. The estimated parameters are qualitatively similar to those in Wang and Tsai (2009). However, the sample fraction analyzed in our model becomes 19.95%, much smaller than the fraction $n_0/n = 54.38\%$ used in the linear model of Wang and Tsai (2009).

Now, we answer the question posed at the beginning of this section by combining the quantitative results from our proposed approach and the economic implication of the covariates. Specifically, there are two possible reasons for the very different performances of these companies. The first one is external due to the fast development of the Chinese economy that makes the capital market extremely volatile and leads to intense competition among many enterprises, particularly for those in the same sector. Particularly, the powerful companies may suffer much less than the weak ones and can keep the return on equity stable. That is, in some sense, the performance of a company partially depends on the economic indicators of the linear component of our model. The second reason is internal caused by factors such as the leverage level (LEV) discussed in this application. It is known that the use of LEV actually makes an investment using borrowed capitals or debts, giving rise to results that may multiply potential return and downside risk simultaneously. Consequently, for a company with high leverage level, its share price can be pushed up quickly in the short run and also perhaps plummet once the high leverage can not sustain, resulting in its withdrawal from the stock market. These two scenarios commonly happen together due to the possible discontinuity, instability and noncompliance of leverage. Thus, companies with higher leverage level tend to have stronger return volatility than those with lower levels, and companies with large negative ROE often face significant financial risks and tend to fall, particularly during the Asian financial crisis from 1997 to 1998.

Finally, as an extended analysis of this data set, we consider two scenarios, in the first of which multiple nonparametric functions are included in our model, while in the second two-way interactions are examined. For the former, we take ASSET, LEV and ATR as nonparametric variables by fitting $\eta_1(\text{ASSET}), \eta_2(\text{LEV})$ and $\eta_3(\text{ATR})$ respectively. We keep PM and SGR in the parametric component since both of them have relatively large standard deviations and outliers that may lead to singularity in computation in local linear estimation. The fitted curves and 95% confidence bands are displayed in Figure 5 while the parametric estimates are presented in the lower panel of Table 4, from which we can get similar conclusions as those in Figure 4 and Table 3 respectively. For the latter, we focus only on including pairwise interactions of the main effects and those between the main effects and LEV. A closer look at this data finds that many cross-terms have very high correlation with the main effects of ASSET, PM, SGR and ATR. After removing any interaction with a correlation coefficient larger than 0.87, we apply backward elimination to further identify significant interactions. It is found that the identified terms are ASSET*LEV and ATR*LEV. The results are reasonable. The ASSET of a company can have significant effect on return on equity due to the influence of LEV, while the high asset turnover ratio of a company may further promote profit based on the influence of LEV. What is unexpected is that, the effects of SGR and ATR are now negative due to the presence of ASSET*LEV and ATR*LEV. In our view, it may reflect an important issue in reality. Specifically, the increase of SGR and ATR can lead to a smaller tail index value and then company will undertake greater risk of extreme events based on the model in equation (1). This happens to be the financial risk accompanied by the use of LEV on the debt and ASSET of a company. We remark that the

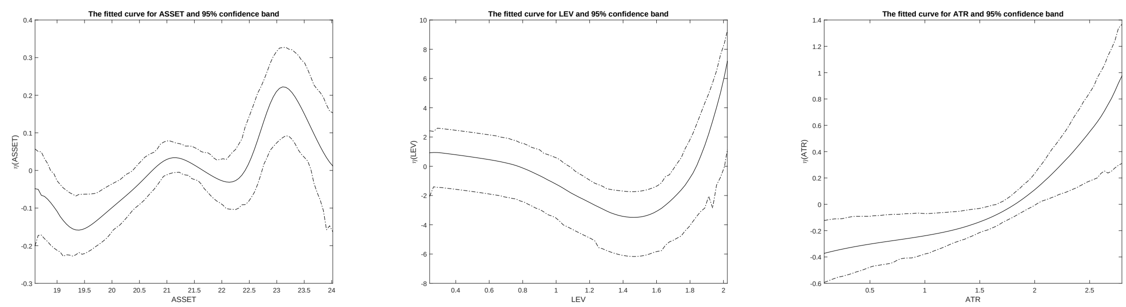

Figure 5: Fitted curves (real line) and the 95% bootstrap confidence bands (dashed line) for η_1(ASSET) in left panel, η_2(LEV) in middle panel and η_3(ATR) in right panel respectively.

Table 4: Tail index estimates (estimate) of parameters in ROE data analysis and the corresponding empirical standard deviation (std), t statistics value(t-stat) and probability of obtaining t-test results (p-value).

covariate	estimate	std	t-stat.	p-value
	Interaction effects			
ASSET	0.0531	0.0183	2.9342	0.0033
PM	0.0405	0.0256	1.4106	0.1584
SGR	-0.0279	0.0112	2.3544	0.0186
ATR	-0.2727	0.0449	6.0363	0.0000
ASSET*LEV	-0.2106	0.0809	2.6977	0.0070
ATR*LEV	0.2834	0.0519	5.3746	0.0000
	Multiple functions			
SGR	0.0505	0.0334	1.2801	0.2005
ATR	0.0141	0.0225	0.4991	0.6177

fitted curve of η(LEV) is similar to that in Figure 4 and thus is not displayed to save space.

Example 3. (Alps meteorology data)

Global warming has gained great concern in scientific community for many years. But does it increase public concern about climate change? Bergquist and Warshaw (2019) concluded that warming climate is unlikely to yield a consensus in the mass public about the threat posed by climate change. We attempt to offer some analysis on this issue by examining the Alps meteorology data. It was reported that the snowing season in Alps reduced 38 days during 1960-2017, which is not just due to the change in temperature but reflects the sharp warming in the climate. It is well-known that Alps reaches from Austria and Slovenia in the east, through Italy, Switzerland, Liechtenstein and Germany to France in the west and brings extreme climate features to these countries. The extreme variation of temperature in Alps may be one of the most effects of global climate change.

We collect 1440 monthly temperature (TEMP) observations from 10 weather stations in Austria in the winter time (calculated by 3 months) during 1970-2017 and analyze it using our proposed method in this article. Specifically, we explore how the significant temperature change depends on hours on sunshine (SUN), total precipitations (RAIN), latitude of weather stations (LATITUDE), year (TIME=1, . . . , 48) and the north Atlantic oscillation index (NAO) on the basis of surface sea-level pressure difference between the Subtropical (Azores) High and the Subpolar (Iceland) Low. Strong positive phases of the NAO tend to be associated with above-normal temperatures in the eastern United States and across northern Europe while below-normal temperatures in Greenland and oftentimes across southern Europe and the Middle East (Jones et al., 1997), one can see http://www.ncdc.noaa.gov/teleconnections/nao/ for more details. Thus, the NAO index is related to the climate of Northern Hemisphere atmosphere and plays a important role in the wintertime of Europe, Mediterranean, parts of the Middle East and eastern North America when exerting a strong control on the climate changes (Hurrell et al, 2003; Osborn, 2011).

In the data set collected from http://www.cru.uea.ac.uk/, the NAO index has continuous observations while the other covariates are discrete integer values. Thus it is natural to model the effect of NAO as a nonlinear function while leaving the other variables as parametric in modeling the tail index. In this application, the optimal threshold is found to be $w_n = 13$, leaving us with $n_0 = 888$ samples. That is, temperatures higher than 13 and lower than -13 are regarded as extreme ones in the Alps area by our model. For the parametric component, the estimates and related statistics are reported in Table 5. We see that both RAIN and TIME have positive though nonsignificant effects on the tail index. On the other hand, the duration of sunshine and the latitude of the weather station significantly promote extreme temperature change based on our analysis. For the covariate SUN, it makes sense because the longer the time of sunshine is, the more likely extreme temperature is. For LATITUDE, our analysis implies that the temperature is sensitive to the latitude of a place.

For the nonparametric component, the fitted curve of the nonlinear function η(NAO) and the 95% confidence band using bootstrap method are displayed in the right panel of

Table 5: Estimated of parameters in Alps data analysis and the corresponding empirical standard deviation (std), t statistics value (t-stat) and p-value.

covariate	estimate	std	t-stat.	p-value
RAIN	0.0422	0.0242	1.6769	0.0936
SUN	-0.0570	0.0263	2.1723	0.0298
LATITUDE	-0.0393	0.0184	2.1979	0.0280
TIME	0.0257	0.0214	1.2179	0.2233

Figure 4. We observe that the NAO index has a nonlinear effect on the tail index that may have led to great changes of the climates in Austria. Indeed, pronounced climate changes have exactly since the 1970s, not just in Austria, which include rapid loss of Arctic sea ice, large-scale climate warming and increased tropical storm activities. The prominent upward trend of NAO from the 1950s to the 1990s caused large regional changes in air temperature, precipitation, wind and storminess, with accompanying impacts on marine and terrestrial ecosystems, and contributed to the accelerated rise in global mean surface temperature (Thompson et al., 2000). More recently, Delworth et al. (2016) showed that the multidecadal variations of the north Atlantic oscillation induces poleward ocean heat transport in the Atlantic extending to the Arctic, and promotes the rapid loss of Arctic sea ice, Northern Hemisphere warming, and changing Atlantic tropical storm activity, especially in the late 1990s and early 2000s.

Finally, we have checked possible correlations in the data and found that there is no spatial correlation but significant temporal correlation. Thus, we decided to use the partially linear dynamic panel data model (Baltagi and Li, 2002; Su and Zhang, 2016) to fit the data again. The resulting model has 11 significant autoregressive lagged orders. The scatter plot in the right panel of Figure 6 below further confirms the existence of the temporal dependence with the correlation coefficient of the residuals becoming from 0.8406 after fitting a naive linear model to 0.1348 when a high-order dynamic model is used. Besides, the fitted curve of NAO index in the left panel of Figure 6 shows similar pattern as that presented in the right panel of Figure 4. We must remark that, however, this dynamic model focuses on the mean not the tail of the data, and thus it would provide a complementary understanding to our analysis.

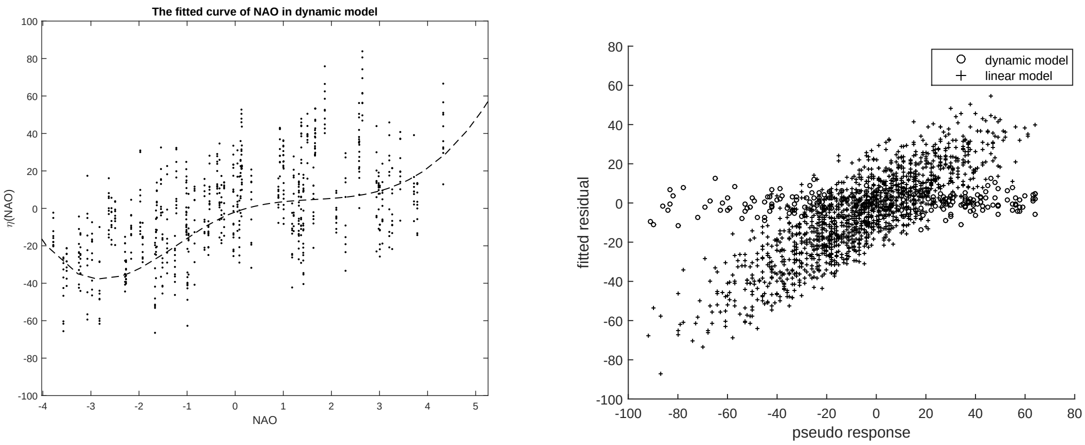

Figure 6: Fitted curves of η(NAO) in left panel and scatter plots of fitted residuals using linear model ('+') and dynamic model ('o') respectively in right panel.

6 Concluding remarks

In this article, we have proposed STIR, a new semiparametric tail index model, for characterizing the dependence of the extreme values of a random variable on predictive variables. We have developed an iterative algorithm to estimate the parametric and nonparametric components and established the asymptotic normality of the resulting estimates. We have shown through synthetic and real examples that our model can provide useful insight into how explanatory variables affect extremeness. We only focused on independent and identically distributed observations in this work. For future research, it is worth considering the dependent case, for example, when the repeated measurements are available (Hsing, 1991; Drees and Rootzén, 2010) and the observations are heterogeneously distributed (Hill, 2010; Einmahl et al., 2016).

Acknowledgements

We thank the Editor, an Associate Editor and two reviewers for their constructive comments that have led to a much improved paper. Li's research was supported by a grant from the National Social Science Fund of China (17BTJ025). You's research was supported by grants from the National Natural Science Found of China (NSFC) (11971291, 11471203) and the Program for the Innovative Research Team of Shanghai University of Finance and Economics

(IRTSHUFE). The work is also partially supported by the Program for Changjiang Scholars and Innovative Research Team in University (IRT13077).

Supporting information

Supplementary materials including lemmas used for establishing the asymptotic properties of resulting estimates and the technical details of the proofs for the main theorems are available online.

References

Baltagi, B. H. and Li, D. (2002). Series estimation of partially linear panel data models with fixed effects. *Ann. Econom. Finan.* **3**(1), 103–116.

Beirlant, J., Bouquiaux, C. and Werker, B. (2006). Semiparametric lower bounds for tail index estimation. *J. Stat. Plan. Infer.* **136**(3), 705–729.

Beirlant, J., Goegebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of Extremes: Theory and Applications.* New York: John Wiley & Sons.

Beirlant, J., Vynckier, P. and Teugels, J. L. (1996). Tail index estimation, pareto quantile plots regression diagnostics. *J. Am. Stat. Assoc.* **91**(436), 1659–1967.

Beran, J. and Schell, D. (2012). On robust tail index estimation. *Comput. Stat. Data Anal.* **56**(11), 3430–3443.

Bergquist, P. and Warshaw, C. (2019). Does global warming increase public concern about climate change? *J. Politc.* **81**(2), 686–691.

Boucheron, S. and Thomas, M. (2015). Tail index estimation, concentration and adaptivity. *Electron. J. Stat.* **9**(2), 2751–2792.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika* **71**(2), 353–360.

Cai, J., Fan, J., Jiang, J. and Zhou, H. (2007). Partially linear hazard regression for multivariate survival data. *J. Am. Stat. Assoc.* **102**(478), 538–551.

Cai, J., Fan, J., Jiang, J. and Zhou, H. (2008). Partially linear hazard regression with varying coefficients for multivariate survival data. *J. R. Stat. Soc.* B **70**(1), 141–158.

Cain, K., Harlow, S., Little, R., Nan, B., Yosef, M., Taffe, J. and Elliott, M. (2011). Bias due to left truncation and left censoring in longitudinal studies of developmental and disease processes. *Am. J. Epidemiol.* **173**(9), 1078–1084.

Chavez-Demoulin,V. and Davison, A.C. (2005). Generalized additive modelling of sample extremes. *J. R. Stat. Soc.* C **54**(1), 207–222.

Chen, H. (1988). Convergence rates for parametric components in a partly linear model. *Ann. Stat.* **16**(1), 136–146.

Chen, L. P. (2019). Pseudo likelihood estimation for the additive hazards model with data subject to left-truncation and right-censoring. *Statistics and Its Interface* **12**(1), 135–148.

Danielssona, J., de Haanc, L., Peng, L. and de Vriesc, C.G. (2001). Using a bootstrap method to choose the sample fraction in tail index estimation. *J. Multivar. Anal.* **76**(2), 226–248.

Davison, A. C. and Ramesh, N. I. (2000). Local likelihood smoothing of sample extremes. *J. R. Stat. Soc.* B **62**(1), 191–208.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with Discussion). *J. R. Stat. Soc.* B **52**(3), 393–442.

Delworth, T. L., Zeng, F., Vecchi, G. A., Yang, X., Zhang, L. and Zhang, R. (2016). The North Atlantic Oscillation as a driver of rapid climate change in the Northern Hemisphere. *Nature Geoscience* **9**(7), 509–512.

Drees, H. and Rootzén, H. (2010). Limit theorems for empirical processes of cluster functionals. *Ann. Stat.* **38**(4), 2145–2186.

Einmahl,J.H.J., de Haan,L. and Zhou, C. (2016). Statistics of heteroscedastic extremes. *J. R. Stat. Soc.* B **78**(1), 31–51.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* London: Chapman and Hall/CRC Press.

Gabaix, X. and Ibragimov, R. (2011). Rank-1/2: a simple way to improve the OLS estimation of tail exponents. *J. Bus. Econom. Stat.* **29**(1), 24–39.

Gomes, M. I., De Haan, L. and Rodrigues, L. H. (2008). Tail index estimation for heavy-tailed models: accommodation of bias in weighted log-excesses. *J. R. Stat. Soc.* B **70**(1), 31–52.

Hall, P. (1982). On some simple estimates of an exponent of regular variation. *J. R. Stat. Soc.* B **44**(1), 37–42.

Härdle, W., Liang, H. and Gao, J. (2012). *Partially Linear Models.* New York: Springer Science & Business Media.

He, X., Zhu, Z. Y. and Fung, W. K. (2002). Estimation in a semiparametric model for longitudinal data with unspecified dependence structure. *Biometrika* **89**(3), 579–590.

Hill, B. M. (1975). A simple general approach to inference about the tail of a distribution. *Ann. Stat.* **3**(5), 1163–1174.

Hill, J. B. (2010). On tail index estimation for dependent, heterogeneous data. *Econom. Theory* **26**(5), 1398–1436.

Howards, P. P., Hertz-Picciotto, I. and Poole, C. (2006). Conditions for bias from differential left truncation. *Am. J. Epidemiol.* **165**(4), 444–452.

Hsing, T. (1991). On tail index estimation using dependent data. *Ann. Stat.* **19**(3), 1547–1569.

Hudgens, M. G. (2005). On nonparametric maximum likelihood estimation with interval censoring and left truncation. *J. R. Stat. Soc.* B **67**(4), 573–587.

Hurrell J.W., Kushnir Y., Ottersen G. and Visbeck M.(eds) (2003). The North Atlantic Oscillation: climate significance and environmental impact. *Geophys. Monogr. Ser* **134**, AGU, Washington, D. C.

Ibragimov M., Ibragimov R. and Kattuman P. (2013). Emerging markets and heavy tails. *J. Bank.* & *Fina.* **37**(7), 2546–2559.

Jia, M., Taufer, E. and Dickson, M. M. (2018). Semi-parametric regression estimation of the tail index. *Electron. J. Stat.* **12**(1), 224–248.

Jones, P. D., Jonsson, T., and Wheeler, D. (1997). Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. *International Journal of Climatology: A Journal of the Royal Meteorological Society* **17**(13), 1433–1450.

Kalbfleisch, J. D. and Prentice, R. L. (2011). *The Statistical Analysis of Failure Time Data.* John Wiley & Sons.

Li, J. and Lee, M. L. T. (2011). Analysis of failure time using threshold regression with semi-parametric varying coefficients. *Stat. Neerla.* **65**(2), 164–182.

Li, J., Zhang, W. and Wu, Z. (2011). Optimal zone for bandwidth selection in semiparametric models. *J. Nonpar. Stat.* **23**(3), 701–717.

Liang, H., Wang, S. J., Robins, J. M. and Carroll, R. J. (2004). Estimation in partially linear models with missing covariates. *J. Am. Stat. Assoc.* **99**(466), 357–367.

Lin, X. and Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *J. Am. Stat. Assoc.* **96**(455), 1045–1056.

Liu, R. and Yang, L. (2010). Spline-backfitted kernel smoothing of additive coefficient model. *Econom. Theory* **26**(1), 29–59.

Ma, Y., Jiang, Y. and Huang, W. (2018). Empirical likelihood based inference for conditional Pareto-type tail index. *Stat.* & *Prob. Lett.* **134**, 114-121.

McElroy, T. and Nagaraja, C. H. (2016). Tail index estimation with a fixed tuning parameter fraction. *J. Stat. Plan. Infer.* **170**, 27–45.

Müller, S. and Rufibach, K. (2009). Smooth tail-index estimation. *J. Stat. Comp. Simul.* **79**(9), 1155–1167.

Nemeth, L. and Zempleni, A. (2017). Regression estimator for the tail index. *arXiv preprint arXiv*:1708.04815v3.

Nicolau, J. and Rodrigues, P. M. (2019). A new regression-based tail index estimator. *Rev. Econom.* & *Stat.* **101**(4), 667–680.

Osborn, T. J. (2011). Winter 2009/2010 temperatures and a record-breaking North Atlantic Oscillation index. *Weather* **66**(1), 19–21.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimator. *Scand. J. Stat.* **9**(2), 65–78.

Schisterman, E. F., Cole, S. R., Ye, A. and Platt, R. W. (2013). Accuracy loss due to selection bias in cohort studies with left truncation. *Paediatr. Perinat. Epidemiol.* **27**(5), 491–502.

Severini, T. A. and Staniswalis, J. G. (1994). Quasi-likelihood estimation in semiparametric mdoels. *J. Am. Stat. Assoc.* **89**(426), 501–511.

Shao, F., Li, J., Ma, S. and Lee, M. L. T. (2014). Semiparametric varying-coefficient model for interval censored data with a cured proportion. *Stat. Med.* **33**(10), 1700–1712.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis.* Chapman and Hall, London.

Speckman, P. (1988). Kernel smoothing in partial linear models. *J. R. Stat. Soc.* B **50**(3), 413–436.

Stone, C. J. (1984). An asymptotically optimal window selection rule for kernel density estimates. *Ann. Stat.* **12**(4), 1285–1297.

Straetmans, S. and Candelon, B. (2013). Long-term asset tail risks in developed and emerging markets. *J. Bank. & Fina.* **37**(6), 1832–1844.

Su, L. and Zhang, Y. (2016). Semiparametric estimation of partially linear dynamic panel data models with fixed effects. *Essays in Honor of Aman Ullah, Advan. Economet.* **36**, 137–204.

Sun, Y., Li, J. and Zhang, W. (2012). Estimation and model selection in a class of semiparametric models for cluster data. *Ann. Instit. Statist. Math.* **64**(4), 835–856.

Thompson, D. W., Wallace, J. M. and Hegerl, G. C. (2000). Annular modes in the extratropical circulation. Part II: Trends. *J. Clim.* **13**(5), 1018–1036.

Tsai, W. Y. (1988). Estimation of the survival function with increasing failure rate based on left truncated and right censored data. *Biometrika* **75**(2), 319–324.

Wang H. and Tsai, C. L. (2009). Tail index regression. *J. Am. Stat. Assoc.* **104**(487), 1233–1240.

Wong, H., Ip, W. and Zhang, R. (2008). Varying-coefficient single-index model. *Comput. Stat. Data Anal.* **52**(3), 1458–1476.

Xia, X., Jiang, B., Li, J. and Zhang, W. (2016). Low-dimensional confounder adjustment and high-dimensional penalized estimation for survival analysis. *Lifetime Data Anal.* **22**(4), 547–569.

Zhang Q., Li D. and Wang H. (2013). A note on tail dependence regression. *J. Multivar. Anal.* **120**, 163–172.